United States Government Memorandum

TO: Magalie Roman Salas
FROM: D. Mark Kennet
SUBJECT: **GEOGRAPHIC DATA AND FORWARD-LOOKING COST MODELLING, Dockets 96-45 and 97-160**
DATE: 23 December, 1997

One of the consequences of the 1996 Telecommunications Policy Act has been to make the implicit universal service subsidy explicit. The FCC is attempting to set the level and allocation of the subsidy using economic principles of forward-looking cost. To this end, three models have been proposed, the Hatfield Model (HM), the Benchmark Cost Proxy Model (BCPM), and the Hybrid Cost Proxy Model (HCPM). In this note, I attempt to lay out a series of logical principles and inferences connecting the intentions of universal service modelling and the work we have done in the HCPM.

> **Principle 1:** *The ideal model for determining "forward-looking" economic cost would be an engineering optimization/planning model that produces as its output the technology mix, network routing, and total cost for any service territory.*
>
> **Principle 2:** *Such a model would require precise data on the geographic location of customers* AND *a map containing the road-connected distances between customers, as well as between each customer and the central office.* Given such data, algorithms exist which can minimize the cost of building any network that will provide service of any given level of quality to the customer base.
>
> **Principle 3:** *At the current state of technology, no map meeting both the location AND distance requirements of Principle 2 exists.* Data are available (with varying degrees of resolution) on customer location, but distances must be approximated using an appropriate metric. All three models currently use the L-1 metric, with the BCPM including a Cartesian distance for part of its feeder plant calculation.
>
> **Principle 4:** *The salient feature of geography for telephone plant planning is the dispersion of customers.* Any good customer location model MUST capture the dispersion that exists in the data, whether the data are at the level of individual customers or at the Census block or higher level. I argue that the BCPM, and, to a lesser extent, the HM, do NOT adequately capture this dispersion given the data they are taking as input. In the case of the BCPM, population is aggregated into four arbitrarily defined microgrids centered on "road-network centroids." Such aggregation can either concentrate or overly disperse population, depending on the location and amount of road network in each quadrant of the BCPM grid. In the HM, the problem is less severe, but I note that once HM clusters are defined the model treats each cluster as a square rather than exploiting the dispersion information contained in their dataset.

To see how the models will treat raw data differently, let us consider an imaginary dataset. To simplify matters, let us suppose that the HCPM grids and BCPM grids have the same size (they do not, even if we were to agree on the appropriate grid size, but that is another matter). The data to consider are in Figure 1, in which customer locations are denoted in red; CB internal points are denoted by stars; CB boundaries are denoted by solid lines.

Figure 2 illustrates what BCPM might do with these data. The green "plus" sign indicates the location of the serving area interface for this grid block; the blue squares indicate the microgrids created by multiplying road distance by 1000. As we see, in this example, the BCPM approach will fairly drastically understate customer dispersion and location. Loop lengths will be incorrect, although average loop length may be approximately correct (average loop length is NOT an adequate summary statistic for this type of analysis).

Figure 3 suggests that when the Hatfield approach is given complete, verified geocoded data, it will perform well -- better than HCPM, illustrated in Figure 5. Note that in Figures 3 and 4 the light dashed line denotes cluster boundaries, while the heavy dashed line denotes the square created to capture the area of the cluster. Dispersion is captured; while the loop plant will be overestimated, the approximation seems reasonable, and I cannot argue that HCPM using CB data will outperform this approach. However, Figure 4 indicates that if none of the customers can be geocoded, the HM "pseudo-geocode" approach does no better than HCPM using Census blocks and may actually be worse, since the model may incur unnecessary investment to handle the two outlier points in the upper left corner.

Figure 5 illustrates the HCPM approach. In this approach, the micrograms are approximately equal in area to the average area of all the CBs in the gridblock. All households and businesses within a CB are assigned to the microgrid containing the CB's internal point; thus, the middle microgrid in the leftmost column shows zero customers, since it has no internal point. This approach will, on average, slightly overstate dispersion but will get distances and hence looplengths approximately right.

Figure 6 illustrates what I would consider the ideal case, in which geocoding affords HCPM CENBLOCK the possibility of using the smallest possible grid size. In this event, only HCPM will model a network that actually fits the customer dispersion as it exists on the map. This is because the HCPM FEEDDIST module actually models plant only to populated microgrids, exploiting any cluster economies that may exist by optimizing the location and number of SAIs within the grid.

In summary, I would submit that the BCPM road network centroid approach introduces unnecessary noise into the modelling effort without any apparent benefit or additional accuracy. HCPM's approach is, I believe, more accurate, while at the same time affording the user the option of introducing geocoded customer locations when they are deemed reliable. The HM geocoding approach -- where it actually uses real, verified geocoded data -- is likely to be somewhat more accurate than the HCPM using Census blocks, but this advantage to the Hatfield model disappears when it is necessary to revert to "pseudo-geocodes."

Finally, I note that while the BCPM may be able to accept geocoded customer locations, it currently offers no particularly advantageous way to treat them; in fact, Indetec has offered no specific mechanism for incorporating geocoded data.

It is worth noting that the foregoing discussion assumes that all data used by the models are equally accessible, equally verifiable, and equally accurate. In fact, such is not the case. HM uses a purely proprietary database purchased from a large mail marketing firm which they have processed through geocoding software. While I am not familiar with the particular software they use to translate street addresses to GIS locations, my experience with some of these packages suggests that their procedure of imputing locations based on house number can lead to significant inaccuracies, particularly in rural areas. For example, in one case with which I am intimately familiar, the rural address 228 Mill Creek Road, Kutztown, PA 19530 appears about one-half mile away from where it is actually located. It is also impossible to verify the address databank itself. There may be significant selection bias (systematic exclusion from the sample) against low-income and rural customers in such a database, but there is no possibility -- certainly within the time frame of this proceeding -- of determining the presence or absence of this bias.

BCPM, on the other hand, uses a proprietary road network database. While the BCPM proponents have stated their intention to make this database available for inspection and verification, it also would appear to be very difficult to verify. Inaccuracies in the lengths of roads within a grid could lead to gross over- or underestimates of the cost necessary to serve the region of interest.

HCPM's data sources all exist either on the public record or are available "off the shelf" from vendors. The specific off-the-shelf vendor data are wire center boundaries, which presumably parties in the proceedings themselves will be able to verify and, if need be, rectify.

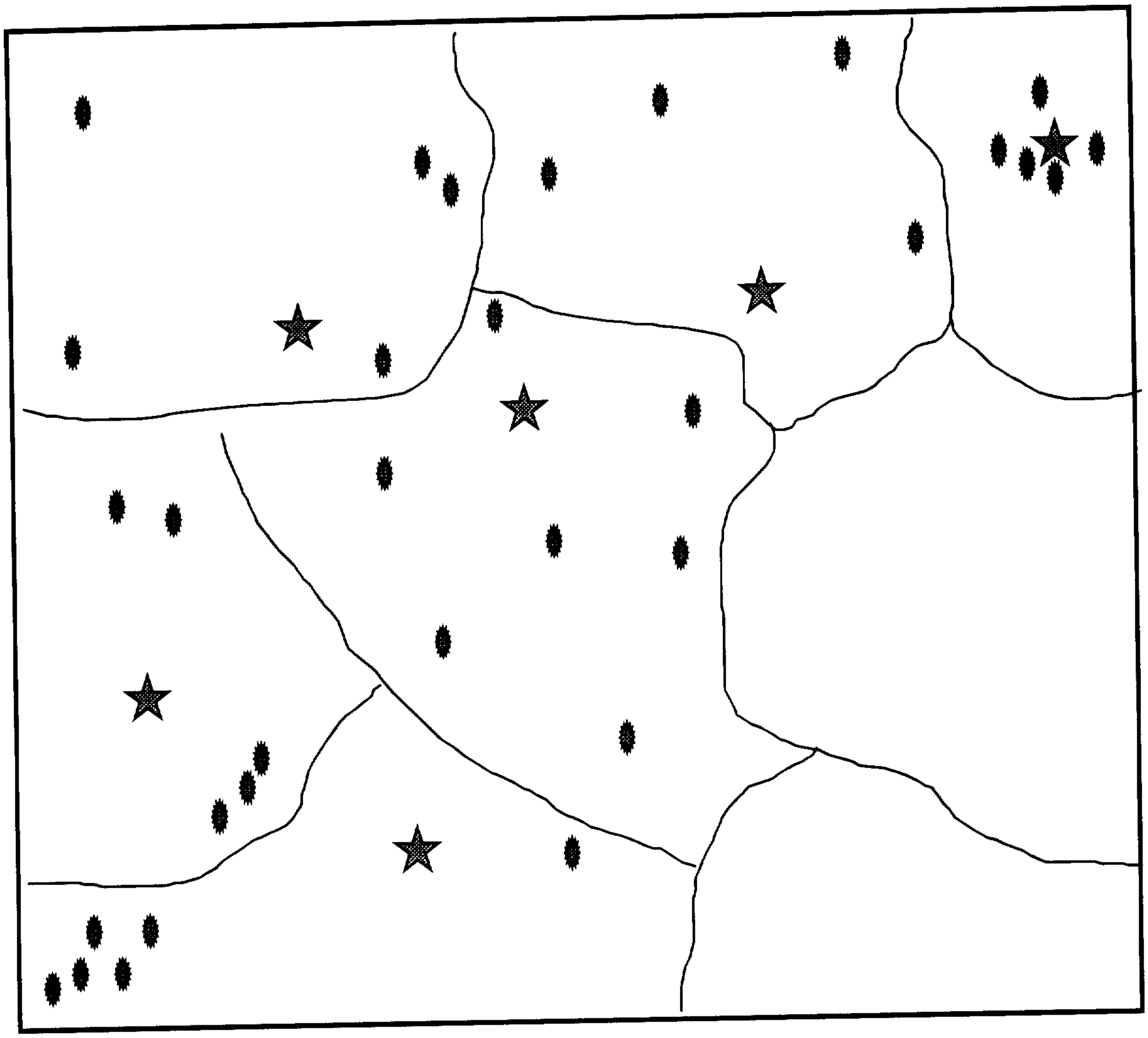

Figure 1

BCPM Approach

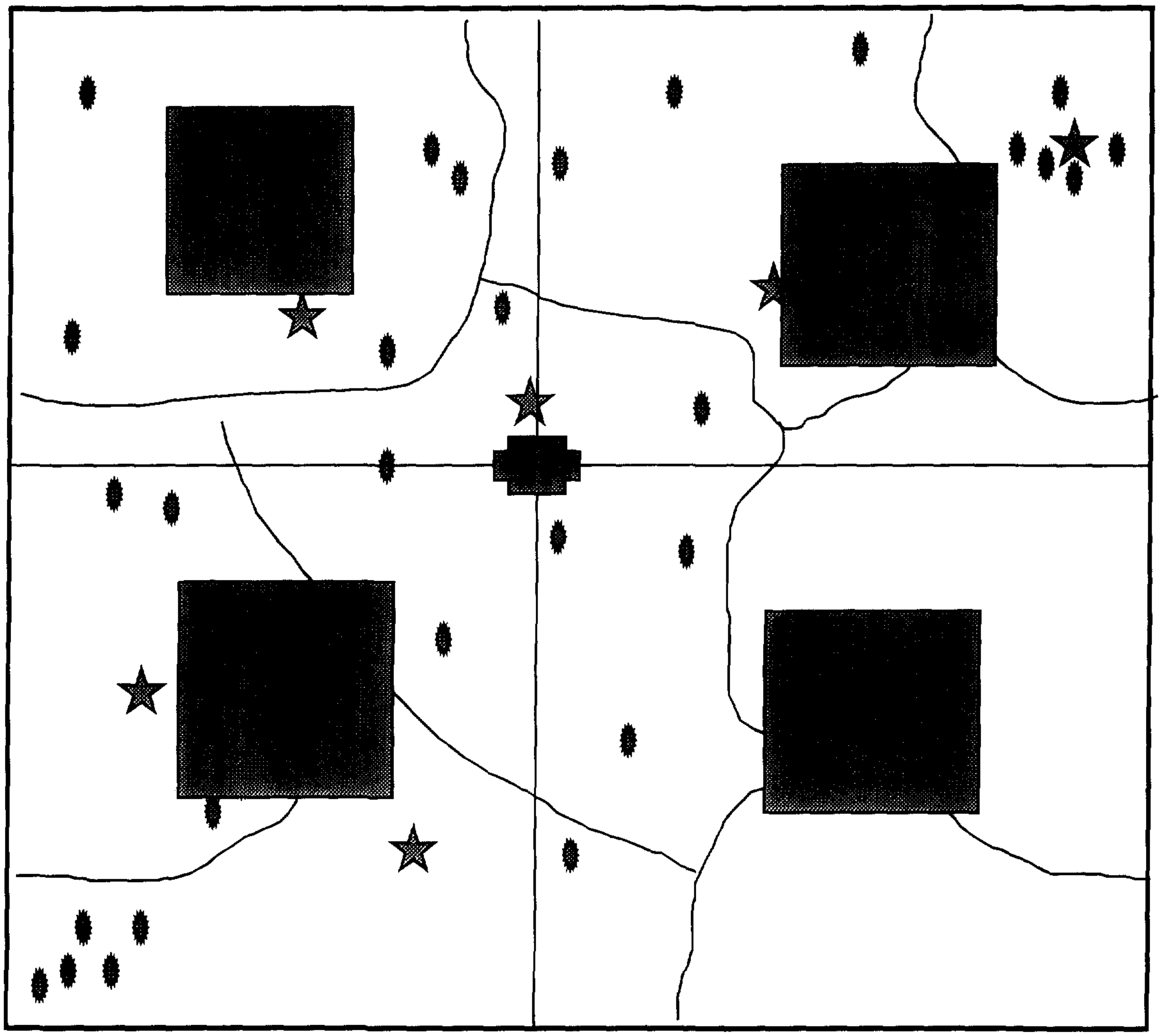

Figure 2

Hatfield Approach when all locations are geocoded

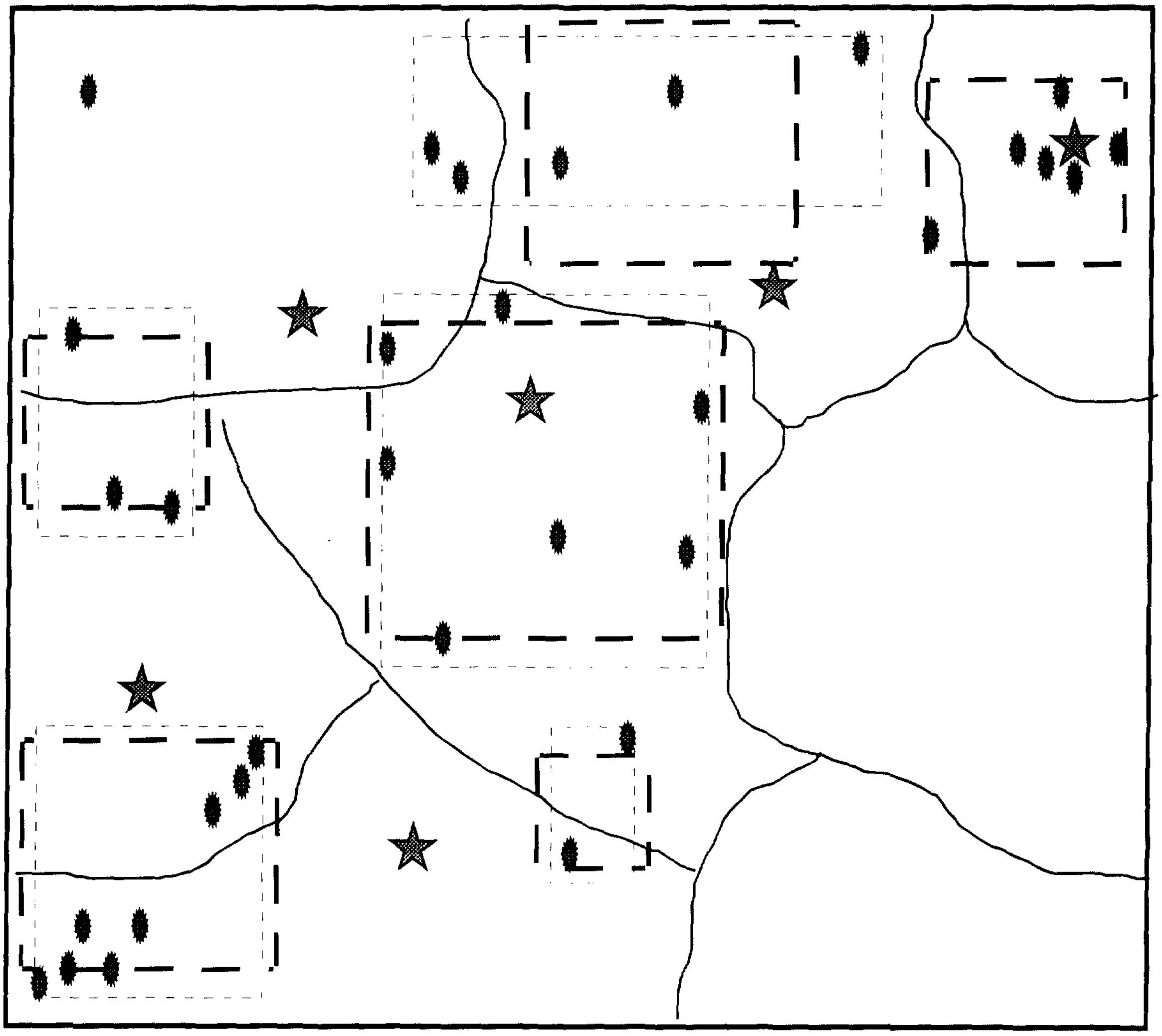

Figure 3

Hatfield results when no customers are geocoded

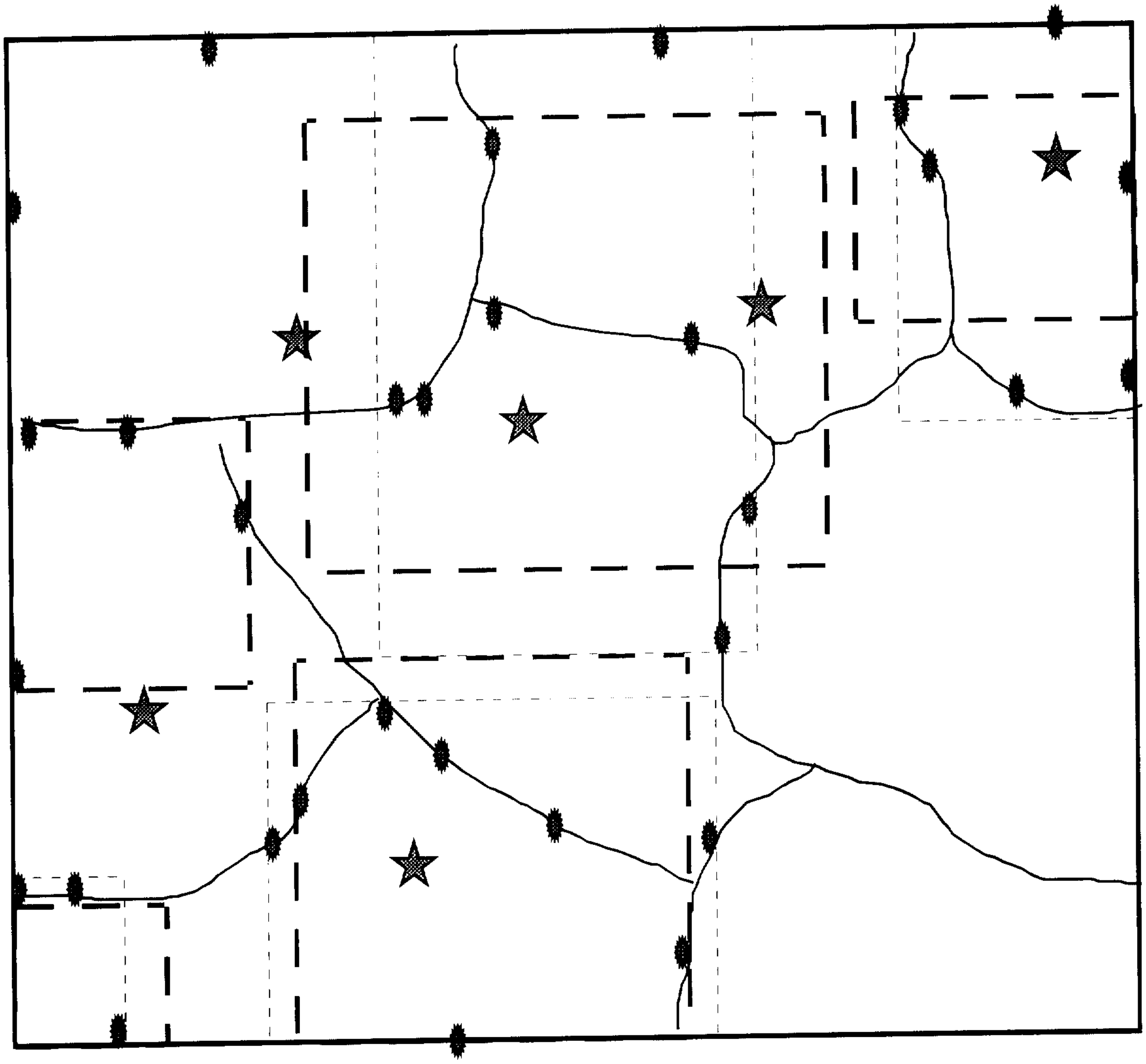

Figure 4

HCPM approach when CBs are used

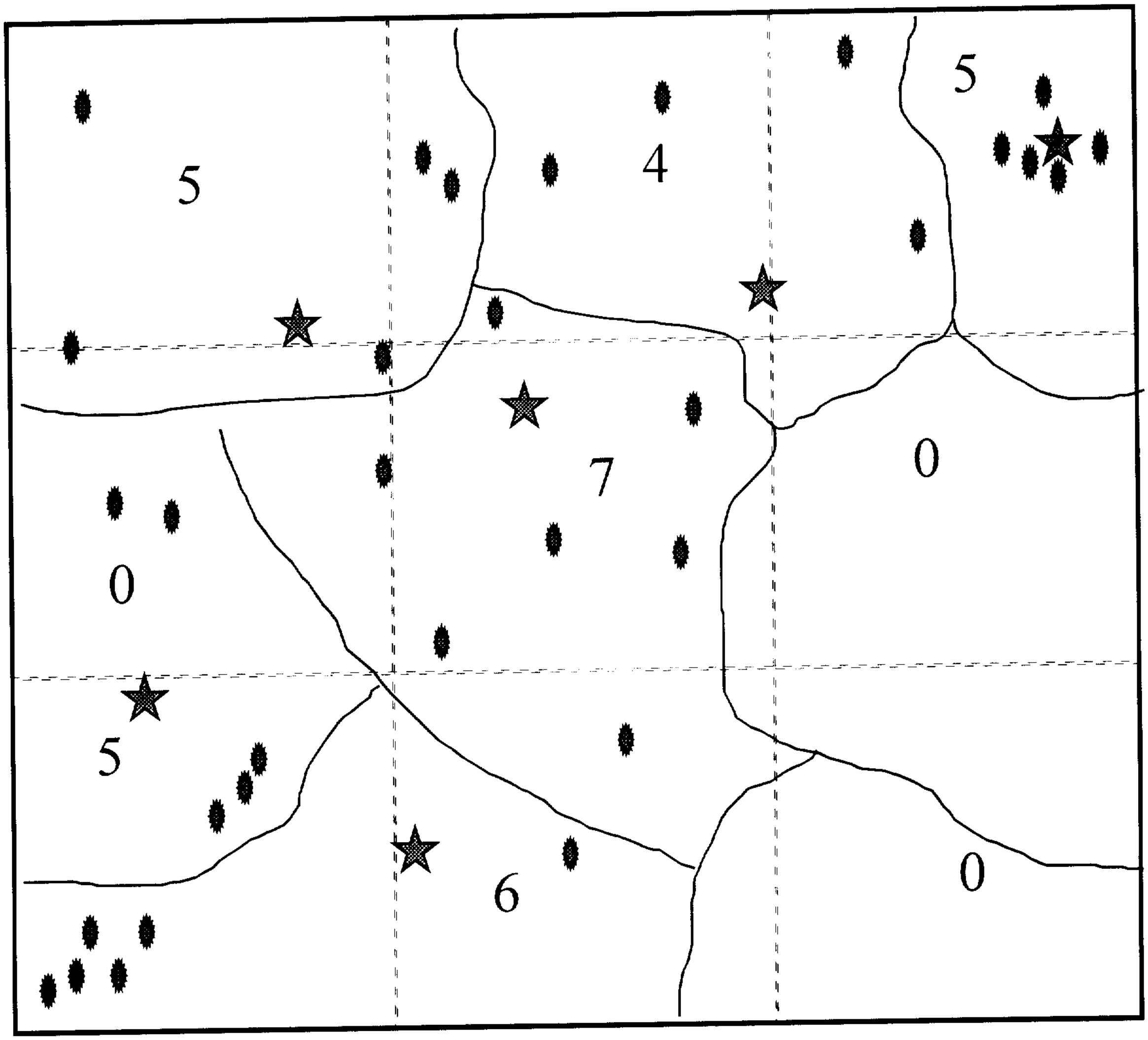

Figure 5

HCPM approach for geocoded points

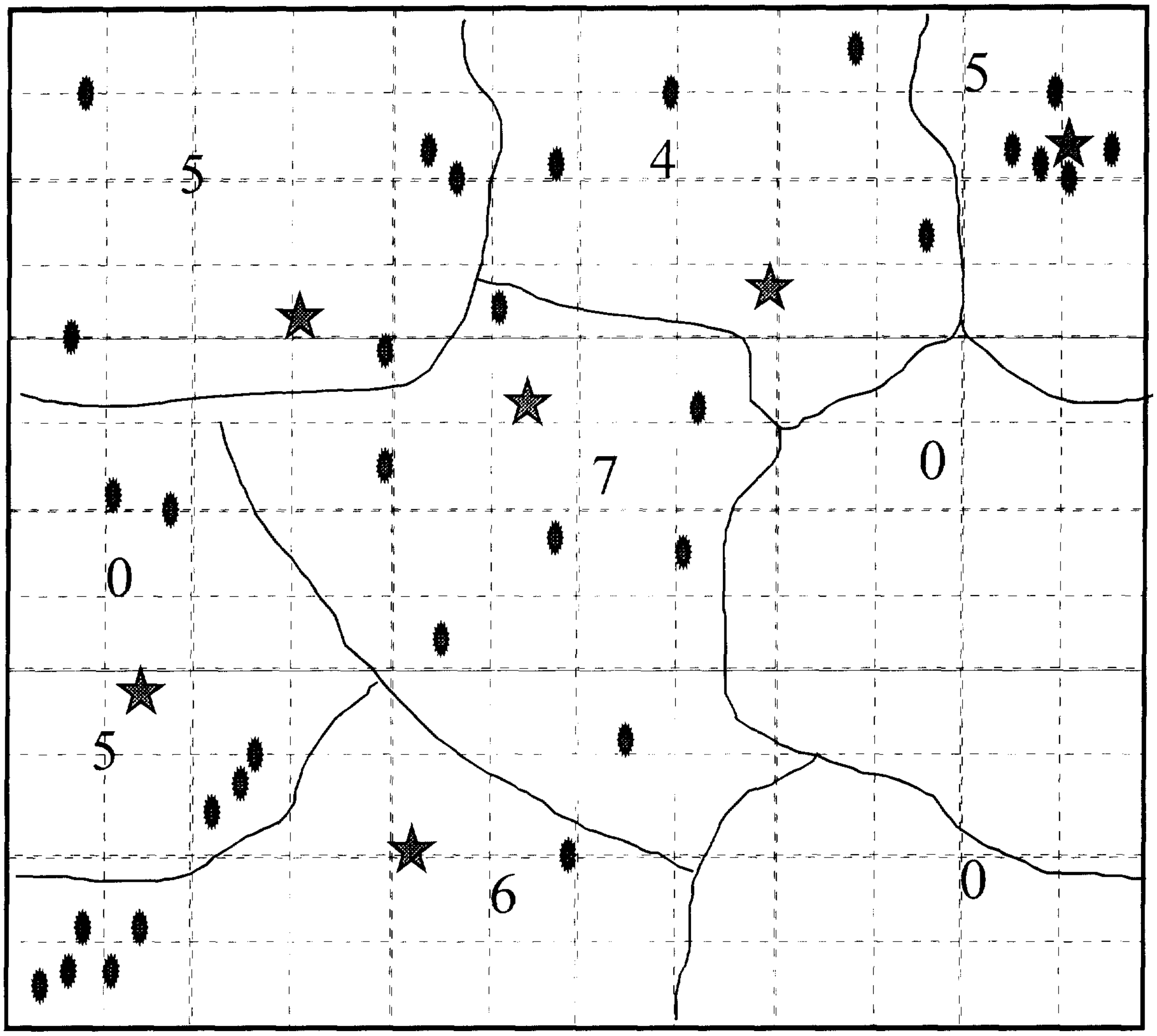

Figure 6